Semiparametric Estimation

Entry for the "New Palgrave Dictionary of Econometrics, Second Edition"

James L. Powell
University of California at Berkeley
October 2006

Abstract

Semiparametric estimation methods are used for models which are partly parametric and partly nonparametric; typically the parametric part is an underlying regression function which is assumed to be linear in the observable explanatory variables, while the nonparametric component involves the distribution of the model's "error terms" that embody the influence of unobservable explanatory variables. The correspondilng estimators of the parametric components may or may not involve explicit nonparametric estimators of unknown features of the error distribution or other unknown functions in the structural model. Semiparametric methods are particulary useful for limited dependent variable models (e.g., the binary response or censored regression models), since fully parametric specifications for those models yield inconsistent estimators if the parametric distribution of the errors is misspecified. The estimation approach for the parametric component varies with the particular form of the parametric relation between the dependent variable, regressors, and error terms, and on the particular identifying restrictions imposed on the conditional distribution of the unobservable error terms given the regressors.

1. Introduction

Semiparametric estimation methods are used to obtain estimators of the parameters of interest – typically the coefficients of an underlying regression function — in an econometric model, without a complete parametric specification of the conditional distribution of the dependent variable given the explanatory variables (regressors). A structural econometric model relates an observable dependent variable y to some observable regressors x, some unknown parameters β, and some unobservable "error term" ε, through some functional form $y = g(x, \beta, \varepsilon)$; in this context, a semiparametric estimation problem does not restrict the distribution of ε (given the regressors) to belong to a parametric family determined by a finite number of unknown parameters, but instead imposes only broad restrictions on the distribution of ε (e.g., independence of ε and x, or symmetry of ε about zero given x) to obtain identification of β and construct consistent estimators of it.

Thus the term "semiparametric estimation" is something of a misnomer; the same estimator can be considered a parametric, semiparametric, or nonparametric estimator depending upon the restrictions imposed upon the economic model. For example, if a random sample of dependent variables $\{y_i\}$ and regressors $\{x_i\}$ are assumed to satisfy a linear regression model $y_i = x_i'\beta + \varepsilon_i$, the classical least squares estimator can be considered a "parametric" estimator of the regression coefficient vector β if the error terms $\{\varepsilon_i\}$ are assumed to be normally distributed and independent of $\{x_i\}$. It could alternatively be considered a "nonparametric estimator" of the best linear predictor coefficients $\beta = [E(x_i x_i')]^{-1}E(x_i y_i)$ if only the weak condition $E(x_i \varepsilon_i)$ is imposed (implying that β is a unique function of the joint distribution of the observations). And the least squares estimator would be "semiparametric" under the intermediate restriction $E(\varepsilon_i|x_i) = 0$, which imposes a parametric (linear) form for the conditional mean $E(y_i|x_i) = x_i'\beta$ of the dependent variable but imposes no further restrictions on the conditional distribution. So the term "semiparametric" is a more suitable adjective for models which are partly (but not completely) parametrically specified than it is for the estimators of those parameters.

Nevertheless, while most econometric estimation methods that do not explicitly specify the likelihood function of the observable data (e.g., least squares, instrumental variables, and generalized method-of-moments estimators) could be considered semiparametric estimators, "semiparametric" is sometimes used to refer to estimators of a finite number of parameters of interest (here, β) that involve explicit nonparametric estimators of unknown nuisance functions (for example, features of the distribution of the errors ε) . Such "semiparametric estimators" use nonparametric estimators of density or regression functions as inputs to second-stage estimators of regression coefficients or similar parameters. Occasionally terms like "semi-nonparametric," "distribution-free," and even "nonparametric" have been used to describe such estimation methods, with the latter terms referring to the treatment of the error terms in an otherwise-parametric structural model.

The primary objective of semiparametric methods is to identify and consistently estimate the unknown parameter of interest β by determining which combinations of structural functions $g(x, \beta, \varepsilon)$ and weak restrictions on the distribution of the errors ε permit this. Given identification and consistent estimation, the next step in the statistical theory is determination of the speed with which the estimator $\hat{\beta}$ converges to its probability limit β. The rate of convergence for estimators for standard parametric problems is the square root of the sample size n, while nonparametric estimators of unknown density and regression functions (with continuously-distributed regressors) generically converge at a slower rate; if a semiparametric estimator can be shown to converge at the parametric rate, i.e., if it is "root-n consistent," then its relative efficiency to a parametric estimator (for a correctly-specified parametric model) will not tend to zero as n increases. For inference, it is also useful to demonstrate the asymptotic (i.e., approximate) normality of the distribution of $\hat{\beta}$ in large samples, so that asymptotic confidence regions and hypothesis tests can be constructed using normal sampling theory. Finally, for problems where existence of root-n consistent, asymptotically-normal semiparametric estimators can be shown, the question of efficient estimation arises. The solution to this question has two parts – determination of the efficiency bound for the semiparametric estimation problem and construction of a feasible estimator that attains that bound.

2. Econometric Applications

In econometrics, most of the attention to semiparametric methods dates from the late 1970s and early 1980s, which saw the development of parametric models for *discrete* and *limited dependent variable (LDV)* models. Unlike the linear regression model, those models are not additive in the underlying error terms, so the validity (specifically, the consistency) of maximum likelihood and related estimation methods depends crucially on the assumed parametric form of the error distribution. As shown for particular examples by Arabmazar and Schmidt (1981, 1982) and Goldberger (1983), failure of the standard assumption of normally-distributed error terms makes the corresponding likelihood-based estimators inconsistent. This is in contrast to the linear regression model, where the maximum likelihood (classical least squares) estimator is consistent under much weaker assumptions than normally (and identically) distributed errors.

Much of the early literature on semiparametric estimation concentrated on a particular limited dependent variable model, the *binary response model*, which arguably presents the most challenging setting for identification and estimation of the underlying regression coefficients. Early examples of semiparametric identification assumptions and estimation methods for this model give a flavor of the approaches used for other econometric models, among them the *censored regression* and *sample selection* models. The discussion here treats only selected assumptions and estimators for these models, and not their numerous variants; more complete surveys of semiparametric models and estimation methods are given by Manski (1989), Powell (1944), Newey (1994), and Pagan and Ullah (1999).

2.1. Semiparametric Binary Response Models

The earliest semiparametric estimation methods in the econometrics literature on LDV models iconcerned the *binary response model*, in which the dependent variable y assumed the values zero or one depending upon the sign of some underlying latent (unobservable) dependent variable y^* which satisfies a linear regression model $y^* = x'\beta + \varepsilon$; that is,

$$y_i = 1\{x_i'\beta - \varepsilon_i > 0\},$$

where "$1\{A\}$" denotes the indicator function of the event A, i.e., it is one if A occurs and is zero otherwise. For a parametric model, in which the errors ε_i are assumed to be independent of x_i and distributed with a known marginal cumulative distribution function $F(\varepsilon)$, the average log-likelihood function takes the form

$$L_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i \ln F(x_i'\beta) + (1-y_i)\ln\left(1 - F(x_i'\beta)\right)\right]$$

for a random sample of size n, and consistency of the corresponding the maximum likelihood estimator $\hat{\beta}_{ML}$ requires correct specification of F unless the regressors satisfy certain restrictions (as discussed by Ruud 1986). When F is unknown, a scale normalization on β is required, and a constant (intercept) term will not be identified no normalization on the location of ε is imposed.

Manski (1975, 1985) proposed a semiparametric alternative, termed the "maximum score" estimator, which defined the estimator to maximize the number of correct matches of the value of y_i with an indicator function $1\{x_i'\beta > 0\}$ of the positivity of the regression function. That is,the maximum score estimator $\hat{\beta}_{MS}$ maximizes the average "score" function

$$S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i \cdot 1\{x_i'\beta > 0\} + (1-y_i)\cdot 1\{x_i'\beta \leq 0\}\right]$$

over β. Unlike the maximum likelihood estimator $\hat{\beta}_{ML}$, consistency of $\hat{\beta}_{MS}$ requires only that the median of the error terms was zero given the regressors, i.e., the conditional cumulative $F(\varepsilon|x)$ of ε_i given $x_i = x$ had $F(\lambda|x) > 1/2$ when $\dot{\lambda} > 0$, and $F(\lambda|x) < 1/2$ when $\lambda < 0$. However, the estimation approach is generally not root-n consistent (as shown by Chamberlain 1986). A variant of the maximum score estimator, proposed by Horowitz (1992), essentially "smoothed" the indicator functions for positivity of $x_i'\beta$ in the minimand $S_n(\beta)$ using a continuous approximation to it, similar to the smoothing used in nonparametric kernel estimators of regression and density functions. The rate of convergence of the resulting "smoothed maximum score" estimator can be made arbitrarily close to the root-n rate if the distribution of the regressors is sufficiently smooth.

To obtain root-n consistent estimators of the unknown β, the assumption on the error term ε can be strenghtened to independence of ε and x. Han (1987) proposed an alternative to the maximum score estimator, termed the "maximum rank correlation" estimator, which compared the sign of the difference $y_i - y_j$ of the dependent variable to the corresponding difference $(x_i - x_j)'\beta$ in the regression functions across all distinct pairs of observations i and j. The estimator $\hat{\beta}_{MRC}$ maximizes

$$M_n(\beta) = \left(\begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(y_i - y_j) \cdot sgn\left((x_i - x_j)'\beta\right)$$

⇔over β, where $sgn(u) \equiv 1\{u > 0\} - 1\{u < 0\}$. The rationale for this estimator is based upon the monotonicity of $\Pr\{y_i = 1|x\} = F(x_i'\beta)$ in $x_i'\beta$, so that, given $y_i \neq y_j$, $\Pr\{y_i > y_j|x_i, x_j\}$ exceeds 1/2 when $x_i'\beta > x_j'\beta$. Han's article gave conditions under which $\hat{\beta}_{MRC}$ was shown to be consistent, and Sherman (1993) showed that this estimator was root-n consistent and asymptotically normal.

An alternative estimation approach for β under the assumption of independence of u and x combine estimation of the parameter vector β with nonparametric estimation of the unknown distribution function F. Cosslett (1983) proposed a "nonparametric maximum likelihood" estimator $\hat{\beta}_{NPML}$ obtained by simultaneously maximizing the likelihood function $L_n(\beta) = L_n(\beta; F)$ over both β and F, where the latter function is restricted to be nondecreasing with values in the unit interval. While consistency of this estimator could be established, its rate of convergence could not. An alternative estimation method, proposed by Klein and Spady (1993), used kernel regression methods to estimate the unknown distribution function F in the likelihood function. The resulting estimator was shown to be root-n consistent and asymptotically normally distributed under additional regularity conditions; furthermore, the estimator was shown to achieve the semiparametric efficiency bound for this problem, i.e., its asymptotic covariance matrix is the smallest possible among regular estimators of β which impose only the independence restriction between x and u.

Still other estimators for β when u and x are independent exploit the *single index regression* structure of this model, since the conditional expectation of y_i given x_i only depends upon the "single index" $x_i'\beta$:

$$E[y_i|x_i] \equiv g(x_i) = F(x_i'\beta).$$

If the vector of regressors x_i is continuously distributed with joint density function $f_X(x)$ which is continuous for all x, Stoker (1986) noted that the vector of slope parameters β is proportional to the expectation of the derivative of $g(x)$,

$$E\left[\frac{\partial g(x_i)}{\partial x}\right] = E[F'(x_i'\beta)] \cdot \beta.$$

Using integration-by-parts, this "average derivative" can turn be expressed as the expected value of the product of $-y_i$ and the derivative of the logarithm of the density f_X of the regressors,

$$E\left[\frac{\partial g(x_i)}{\partial x}\right] = -E\left[y_i \frac{\partial \log\left[f_X(x_i)\right]}{\partial x}\right].$$

Härdle and Stoker (1989) proposed a semiparametric estimator of this representation of β (up to scale) using nonparametric (kernel) estimators of f_X and its gradient, while Powell, Stock, and Stoker (1989) constructed a similar estimator of the "density-weighted average derivative"

$$\begin{aligned} E\left[f_X(x_i)\frac{\partial g(x_i)}{\partial x}\right] &= E[f_X(x_i)F'(x_i'\beta)] \cdot \beta \\ &= -2E\left[y_i \frac{\partial f_X(x_i)}{\partial x}\right], \end{aligned}$$

which is also proportional to β under the single index restriction.

Though the motivation given here was based upon the binary response model under independence of the errors and regressors, the average derivative and weighted average derivative estimators apply to other models with a single index structure, e.g., any *transformation model* with

$$y_i = T(x_i'\beta + \varepsilon_i),$$

for T a nondegenerate function (possibly unknown) and with ε_i continuously distributed and independent of x_i. The same is trure for the "single index regression" estimator proposed by Ichimura (1993), defined to minimize

$$R_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{F}(x_i'\beta;\beta)\right)^2 t_n(x_i);$$

in this expression, $\hat{F}(u;\beta)$ represents a nonparametric regression estimator of $E[y_i|x_i'\beta = u]$ and $t_n(x_i)$ represents a "trimming" term which is zero whenever x_i lies outside a set for which F is sufficiently precisely estimated. Unlike the average derivative $\hat{\beta}_{AD}$ and weighted average derivative $\hat{\beta}_{WAD}$ estimators, which require the regressors to be jointly continuously distributed, root-n consistency and asymptotic normality of the single index regression estimator $\hat{\beta}_{SIR}$ requires only that $x_i'\beta$ has a continuous distribution, so that some of the regressors can be discrete. The criterion function $R_n(\beta)$ is the nonlinear least squares analogue of the maximand for the Klein and Spady (1993) estimator (which also involved a similar trimming term $t_n(x_i)$). The asymptotic covariance matrices for both estimators have the same general form as the corresponding nonlinear least squares and maximum likelihood estimators with F known, except for the replacement of the cross product of the regressors x_i with the cross product of $x_i - E[x_i|x_i'\beta]$, adjusting the asymptotic covariance matrices upward to account for the nonparametric estimation of the unknown function F.

The problem of consistent estimation of β in binary response models is compounded for *panel data* models with *fixed effects* (i.e., individual-specific intercept terms), written as

$$y_{it} = 1\{x_{it}'\beta + \alpha_i - \varepsilon_{it} > 0\}$$

for individuals i ranging from 1 to n and time periods t from 1 to T. For this model, even if the distribution function F of the error terms ε_{it} is known, the maximum likelihood estimators of β and the fixed effects $\{\alpha_i\}$ will generally be inconsistent if the number of time periods T is fixed as N increases. A consistent semiparametric estimation strategy using a variant of the maximum rank correlation estimator was proposed by Manski (1987); for the special case $T = 2$ (i.e., two time periods), the estimator $\hat{\beta}_{BPD}$ can be defined as the maximizer of the criterion

$$P_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} sgn(y_{i2} - y_{i1}) \cdot sgn\left((x_{i2} - x_{i1})'\beta\right),$$

which is analogous to $M_n(\beta)$, except that the differencing is across time periods rather than across individuals. While consistency of $\hat{\beta}_{BPD}$ was established under weak conditions on the error terms, it is not possible to obtain a root-n consistent estimator unless the errors are logistic (Chamberlain 1993) or other restrictive assumptions (e.g., independence of the fixed effect α_i and the regressors x_{it}, or the conditions in Honoré and Lewbel 2002 or Lee 1999) are imposed.

2.2. Other Semiparametric Econometric Models

Many of the identifying assumptions imposed on semiparametric binary response models give identification and yield consistent estimators for other limited dependent variable models, though these models can sometimes be identified and consistently estimated using assumptions that are uninformative for binary response. Consider, for example, the censored regression model, in which the dependent variable y_i satisfies a linear regression model if it is nonnegative, and is zero otherwise:

$$y_i = \max\{0, x_i'\beta + \varepsilon_i\}.$$

For this model, as for the binary response model, the dependent variable is a monotonic function of the error term ε_i; since monotone transformations by definition preserve orderings, the median (or any other percentile) of this monotonic transformation of ε_i is the monotonic transformation evaluated at the median. Thus the assumption that the errors ε_i have conditional median zero given x_i implies that the conditional median of y_i given x_i takes the form $\max\{0, x_i'\beta\}$, depending only on the unknown coefficients β and not on the shape of the distribution of ε_i. Using this fact, and the characterization of medians as minimizers of a least absolute deviations criterion, Powell (1984) proposed estimation of the unknown β vector by the minimizer $\hat{\beta}_{CLAD}$ of

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} |y_i - \max\{0, x_i'\beta\}|$$

for this model; it is analogous to the maximum score estimator $\hat{\beta}_{MS}$ for the binary response model, which can be defined as the minimizer of the sample average absolute deviation of y_i from its conditional median function $1\{x_i'\beta > 0\}$ for binary response with median zero errors. (The maximum rank correlation estimator $\hat{\beta}_{MRC}$ and binary panel data estimator $\hat{\beta}_{BPD}$ can also be expressed as solutions to least absolute deviations problems.) Unlike $\hat{\beta}_{MS}$, though, the censored median estimator $\hat{\beta}_{CLAD}$ is root-n consistent and asymptotically normally distributed under weak regularity conditions, without need for a scale normalization. An alternative estimator for this model, which involved a nonparametric estimator of the probability that y_i equals zero given x_i, was proposed by Buchinsky and Hahn (1997).

A stronger restriction on the error distribution is conditional symmetry about zero given the regressors; while this restriction is no more informative than the implied zero median restriction for binary response, it yields different identification approaches for censored regression. Specifically, the "symmetrically censored" residual

$$
\begin{aligned}
u_i(\beta) &\equiv \min\{y_i - x_i'\beta, x_i'\beta\} \\
&= \min\{\max\{-x_i'\beta, \varepsilon_i\}, x_i'\beta\}
\end{aligned}
$$

is an even function of ε_i when the regression function $x_i'\beta$ is positive, and thus is itself conditionally symmetric about zero. This implies a population moment restriction

$$
0 = E\left[1\{x_i'\beta > 0\}\psi(\tilde{u}_i(\beta)) \cdot x_i\right],
$$

for $\psi(u) = -\psi(-u)$ an odd function of its argument. Powell (1986) proposed a "symmetrically censored least squares" estimator of β based upon the this restriction with $\psi(u) = u$; like the censored median estimator $\hat{\beta}_{CLAD}$ – which exploits the same moment condition with $\psi(u) = sgn(u)$ – the estimator $\hat{\beta}_{SCLS}$ is root-n consistent and asymptotically normally distributed under weak assumptions. Neither estimator involves explicit nonparametric estimation of the error distribution, a feature shared by the maximum score estimator $\hat{\beta}_{MS}$ and its relatives $\hat{\beta}_{MRC}$ and $\hat{\beta}_{BPD}$ for binary response.

As for the binary response model or most limited dependent variable models, consistent estimation of slope coefficients using panel data with fixed effects is challenging, with maximum likelihood estimators for β being inconsistent when the number of time periods is fixed and the number estimated fixed effects increases. For the special case $T = 2$, writing

$$
y_{it} = \max\{0, x_{it}'\beta + \alpha_i + \varepsilon_{it}\},
$$

Honoré (1992) noted that the difference in "identically trimmed" residuals

$$
\begin{aligned}
\tilde{u}_i(\beta) &= \max\{-x_{i1}'\beta, y_{i2} - x_{i2}'\beta\} - \max\{-x_{i2}'\beta, y_{i1} - x_{i1}'\beta\} \\
&= \max\{-x_{i1}'\beta, -x_{i2}'\beta, \alpha_i + \varepsilon_{i2}\} - \max\{-x_{i1}'\beta, -x_{i2}'\beta, \alpha_i + \varepsilon_{i2}\}
\end{aligned}
$$

would be symmetrically distributed about zero if the error terms ε_{i1} and ε_{i2} were identically distributed given x_{i1} and x_{i2} and value of the fixed effect α_i. This implies population moment conditions of the form

$$0 = E\left[\psi(\tilde{u}_i(\beta)) \cdot (x_{2i} - x_{i1})\right],$$

again with $\psi(u)$ an odd function of its argument. Setting $\psi(u) = sgn(u)$ and $\psi(u) = u$ yields root-n consistent and asymptotically normal estimators which are similar to the censored least absolute deviations estimator $\hat{\beta}_{CLAD}$ and symmetrically-censored least squares estimator $\hat{\beta}_{SCLS}$, respectively.

Other estimation approaches for censored regression involve explicit nonparametric estimation of features of the distribution of the error terms, which is common for other semiparametric econometric models. One such model is the *semiparametric regression* (or *semilinear regression*) model, for which some regressors enter linearly while others enter nonparametrically. The model can be written algebraically as

$$\begin{aligned} y_i &= x_i'\beta + \lambda(w_i) + \varepsilon_i \\ &\equiv x_i'\beta + u_i, \end{aligned}$$

where the error terms ε_i are restricted to satisfy $E[\varepsilon_i|x_i, w_i] = 0$, or, equivalently,$E[u_i|x_i, w_i] = E[u_i|w_i] \equiv \lambda(w_i)$; the regressors x_i and w_i thus enter parametrically (linearly) or nonparametrically in the conditional mean of y_i. Robinson (1988) exploited the fact that

$$y_i - E[y_i|w_i] = (x_i - E[x_i|w_i])'\,\beta + \varepsilon_i$$

to construct a root-n consistent, asymptotically-normal estimator of β by applying least squares estimation to this equation, replacing the unknown quantities $E[y_i|w_i]$ and $E[x_i|w_i]$ by nonparametric (kernel) estimators. For the parameters β to be identified for this model, the covariance matrix of the "residual regressors" must be nonsingular, ruling out functional dependence of x_i on w_i.

Though the semilinear regression model is not itself a limited dependent variable model, it arises as a consequence of "selectivity bias" in a bivariate limited dependent variable model, the *censored selection* model, in which a linear latent "outcome" variable $y_i^* = x_i'\beta + \varepsilon_i$ is observed only if some related binary "selection" variable d_i equals one:

$$
\begin{aligned}
d_i &= 1\{w_i'\delta - \eta_i > 0\}, \\
y_i &= d_i \cdot (x_i'\beta + \varepsilon_i),
\end{aligned}
$$

where the regressors x_i and w_i are observed and the unobserved error terms η_i and ε_i need not be mutually independent. Heckman (1979) showed that, for the uncensored ($d_i = 1$) subsample from this model, the dependent variable satisfied a semilinear regression model, since

$$
E[y_i|d_i = 1, x_i, w_i] = x_i'\beta + \lambda(w_i'\delta);
$$

when the errors are jointly normal, as Heckman assumed, the function $\lambda(u)$ has a known parametric form, but is nonparametric if the error distribution is not in a parametric family. Cosslett (1989) developed a consistent two-step estimator for the regression parameters β in the outcome equation, computing a binary nonparametric maximum likelihood estimator $\hat{\delta}_{NPML}$ of δ in the first step and using a step-function approximation to $\lambda(u)$ in a least squares fit of the outcome equation for the uncensored observations. Ahn and Powell (1993) proposed a root-n consistent two-step estimator of β for the related semilinear model

$$
E[y_i|d_i = 1, x_i, w_i] = x_i'\beta + \lambda^*(p(w_i)),
$$

where the "propensity score" $p(w_i) \equiv E[d_i|w_i]$ is first estimated by a nonparametric regression method; this semilinear model is implied by a generalization of the original censored selection model, replacing the linear form of the regression function $w_i'\delta$ in the selection equation with an unknown function of the regressors w_i.

Some variations of the censored selection model admit other semiparametric identification strategies. For example, if the selection equation equation is censored rather than binary, i.e., if

$$
\begin{aligned}
d_i &= \max\{0, w_i'\delta + \eta_i\}, \\
y_i &= 1\{d_i > 0\} \cdot (x_i'\beta + \varepsilon_i),
\end{aligned}
$$

then Honoré, Kyriazidou, and Udry (1997) construct a root-n consistent two-step estimator of β under the assumption that the errors η_i and ε_i are jointly symmetric about zero given the regressors w_i and x_i, using a symmetrically censored least squares estimator of δ in the first step and exploiting the symmetry of $y_i - x_i'\beta$ about zero given that $0 < d_i < 2w_i'\delta$ in the second step. In contrast, estimation of censored selection models for panel data with fixed effects is no less challenging than for binary panel data models; Kyriazidou (1997) proposes a consistent (but not root-n consistent) two-step estimator for the panel data selection model

$$\begin{aligned} d_{it} &= 1\{w_{it}'\delta + \nu_i - \eta_{it} > 0\}, \\ y_{it} &= d_{it} \cdot (x_{it}'\beta + \alpha_i + \varepsilon_{it}), \end{aligned}$$

using Manski's (1987) binary panel data estimator to estimate δ in the first step and a semilinear regression estimator similar to the Ahn and Powell (1993) approach in the second step.

3. References

Ahn, H. and J.L. Powell (1993), "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3-29.

Arabmazar, A. and P. Schmidt (1981), "Further Evidence on the Robustness of the Tobit Estimator to Heteroscedasticity," *Journal of Econometrics*, 17:253-258.

Arabmazar, A. and P. Schmidt (1982), "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," *Econometrica*, 1055-1063.

Buchinsky, M. and J. Hahn (1998), "An Alternative Estimator for the Censored Quantile Regression Model," Econometrica, 66, 653-672.

Chamberlain, G. (1986), "Asymptotic Efficiency in Semiparametric Models with Censoring," *Journal of Eonometrics*, 32, 189-218.

Chamberlain, G. (1993), "Feedback in Panel Data Models," unpublished manuscript, Department of. Economics, Harvard University.

Cosslett, S.R. (1983), "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51, 765-782.

Cosslett, S.R. (1991), "Distribution-Free Estimator of a Regression Model with Sample Selectivity," in W.A. Barnett, J.L. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (Cambridge University Press).

Goldberger, A.S. (1983), "Abnormal Selection Bias," in S. Karlin, T. Amemiya, and L. Goodman, eds., *Studies in Econometrics, Time Series, and Multivariate Statistics* (Academic Press).

Han, A.K. (1987), "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35, 303-316.

Härdle, W. and T.M. Stoker (1989) "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986-995.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-162.

Honoré, B.E. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-565

Honoré, B.E., E. Kyriazidou, and C. Udry (1997), "Estimation of Type 3 Tobit Models Using Symmetric Trimming and Pairwise Comparisons," *Journal of Econometrics*, 76, 107-128.

Honore B.E. and by A. Lewbel (2002), "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica* 70, 2053-2063.

Horowitz, J.L. (1992), "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505-531.

Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58, 71-120.

Klein, R.W. and R.H. Spady (1993), "An Efficient Semiparametric Estimator for Discrete Choice Models," *Econometrica*, 61, 387-421.

Kyriazidou, E. (1997), "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335-1364.

Lee, M-J. (1999), "A Root-n Consistent Semiparametric Estimator for Related Effect Binary Response Panel Data," *Econometrica*, 60, 533-565.

Manski, C.F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3: 205-228.

Manski, C.F. (1985), "Semiparametric Analysis of Discrete Response, Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 205-228.

Manski, C.F. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357-362.

Manski, C.F. (1989), *Analog Estimation Methods in Econometrics* (Chapman-Hall).

Newey, W.K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62: 1349-1382.

Pagan, A.P. and A. Ullah (1999), *Nonparametric Econometrics* (Cambridge University Press).

Powell, J.L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," J*ournal of Econometrics*, 25, 303-325.

Powell, J.L. (1986), "Symmetrically Trimmed Least Squares Estimation of Tobit Models," *Econometrica*, 54, 1435-1460.

Powell, J.L. (1994), "Estimation of Semiparametric Models," in Engle, R.F. and D.L. McFadden, eds., *Handbook of Econometrics, Vol. 4* (North Holland).

Powell, J.L., J.H. Stock and T.M. Stoker (1989), "Semiparametric Estimation of Weighted Average Derivatives," *Econometrica*, 57, 1403-1430.

Robinson, P. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

Ruud, P. (1986), "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution," *Journal of Econometrics*, 32, 157-187.

Sherman (1993), "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123-137.

Stoker, T.M. (1986), "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461-1481.